DOCUMENT RESUME

ED 198 432 CG 014 966

AUTHOR Araujo, John: Semb, George
TITLE Measuring Concept Levels in Child Psychology Course Material.
PUB DATE Sep 80
NOTE 19p.: Paper presented at the Annual Convention of the American Psychological Association (88th, Montreal, Quebec, Canada, September 1-5, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Abstract Reasoning: Child Development: Cognitive Processes: *Concept Formation: *Concept Teaching: Correlation: *Course Content: Course Evaluation: *Creative Thinking: *Higher Education: Instructional Improvement: Measurement Techniques: Test Theory

ABSTRACT

Levels of learning can probably be measured by questions containing various conceptual dimensions which relate to the amount of thinking or creativity required to complete the questions. A method for measuring concept levels in child development course material was evaluated. Student proctors (N=41) and instructors (N=6) rated 73 short-answer essay questions on four dimensions: difficulty, importance, thinking, and creativity. Objective difficulty data (error frequencies) were also collected. Proctors' and instructors' ratings were significantly different, positively correlated, and significantly reliable. While the objective difficulty data were also found to be reliable, they did not correlate significantly with either difficulty or any of the other three dimensions as rated by proctors or instructors. Results suggest that the method reliably measured concept levels, and could provide useful information about the quality of educational material. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED198432

Measuring Concept Levels in Child Psychology Course Material

John Araujo and George Semb

Department of Human Development

University of Kansas

Lawrence, Kansas 66045

(913) 864-4840

Paper presented in Division 2 of the

American Psychological Association, 1980

CG 014966

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Araujo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

Measuring Concept Levels in Child Psychology Course Materials

Levels of learning can probably be measured by questions containing various conceptual dimensions (Coldeway, 1978). Such dimensions might relate to the amount of reasoning, thinking, or creativity required to complete the question. This premise, however, requires a measurement system for these concepts.

Several educators have proposed methods for concept level measurement (e.g., Bloom, 1954; Gagné 1970; Semb & Spencer, 1976; and Sanchez-Sosa, Semb, & Spencer, note 1). The method proposed by Sanchez-Sosa, Spencer, and Semb (note 1) and Semb and Spencer (1976) has been utilized but not evaluated. This method has also been described by Wolf (1978) as social validation.

A new variation of this system is whether or not it can be used by two different, but relevant audiences. This study investigated this variation in an introductory child psychology course. The audiences were proctors (Keller, 1968) and instructors.

Method

Subjects, setting, and materials. Each of 41 proctors and 6 instructors rated 73 short answer essay questions. These 73 questions represented the entire essay question pool for the 12 units of the course. The proctors rated the questions as part of a special proctoring course. The short answer essay questions were rated in two blocks of 25 and one block of 23 questions, and each block was completed during one of three successive class meetings. The instructors rated

the short answer essay questions during their office hours. Whether they broke the task into blocks similar to the proctors' or completed the task at one time was their decision.

Objective data. Error frequencies of students enrolled in the course provided an objective measure of question difficulty. Ten answers for each of the 73 questions were analyzed for accuracy. Accuracy was expressed as percent correct.

Questionnaire construction. A questionnaire was constructed to measure the conceptual levels of each short answer essay question. The six course instructors jointly discussed and suggested key terms that might describe what questions with higher conceptual levels should contain. For example, should questions be difficult, yet challenging and exciting to answer, and/or should they also require a great deal of preparation time, thinking, and/or creativity? After a consensus prevailed on descriptors of higher conceptual levels, they were stated as questions with associated 7-point semantic differentials (Osgood, Suci, & Tannenbaum, 1957). The instructors again reviewed these questions and made suggestions for their refinement. Finally, four questions were agreed upon as measuring the desired characteristics and/or conceptual levels of the questions: difficulty, importance, thinking, and creativity. A rating of "1" indicated it had the maximal amount of the characteristic, and a "7" indicated it had the least amount of the characteristic being rated.

Analyses. The statistical analyses were divided into three major areas. The first analysis, utilizing a one-way repeated measures

analysis of variance, compared differences within instructors and within proctors. All pairwise post hoc comparisons were performed according to Scheffé (1953). Correlations (Pearson r) were analyzed in the second area. The third area was reliability, of which there were two types. First, the reliability of the ratings provided by the proctors and instructors were analyzed. This analysis was performed as described by Winer (1971, pp. 283-296). The second type of reliability was grading reliability of the 13 short answer essay questions. This was computed by dividing the total number of answers on which the original and reliability grader agreed by the total number of answers checked for reliability.

Results

Differences. Overall, the four dimensions as rated by the proctors were significantly different (see Table 1). Furthermore, all pairwise comparisons were significantly different except for difficulty versus creativity (see Table 2). This was also true for the instructors except that only the difficulty versus thinking comparison was not significant (see Table 3 and 4). Comparisons between proctors and instructors revealed that only thinking and creativity were significantly different from each other (see Table 5).

Correlations. Except for the importance dimension, there were significant positive correlations between the proctor and instructor ratings (see Table 6). The objective difficulty data did not correlate significantly with either the proctors' or the instructors' ratings of difficulty (see Tables 7 and 8).

Reliability. Both the proctor and instructor data for each of the dimensions were found to be significantly reliable (see Table 9). Percent agreement for the 73 short answer essay questions was 80.65%.

Discussion

This study examined the behavioral assessment of four concept levels in university course materials. Although previous researchers and educators have recommended this method's use (Semb & Spencer, 1976; Wolf, 1978; and Sanchez-Sosa, Semb, & Spencer, note 1), data have not been gathered upon which an empirical recommendation might be made. However, the present data provide support for this method of assessing concept levels.

It was found that objective (error frequencies) and subjective difficulty did not correlate. Perhaps, one or both of the assessments were measuring something other than difficulty, and therefore, it is reasonable for these two assessments not to correlate. On the other hand, all three of the remaining dimensions correlated worse than did the subjective difficulty with objective difficulty data. Since both data sets were reliable, the educator might have difficulty in choosing a method for measuring this concept in course materials. Future research might improve this correlation by immediately following the response to a question with its rating. In this way, the requirements for answering the question might have a more accurate and reliable affect upon the rating.

The significant differences between instructor and proctor ratings of thinking and creativity is not surprising. Instructors, presumably,

are more knowledgeable and possess an academic repertoire different from that of the proctors. Thus, instructors would be expected to find the materials as requiring less thinking and creativity than would the proctors.

Among the important findings were the significant positive correlations between the two relevant audiences. Kazdin (1977) originally raised this issue as an important consideration when using social validation, and Gagné (1971) and Merrill (1971) specified it as a criterion for an adequate task analysis. The present data begin a body of evidence suggesting that two groups can reliably rate subjective data.

The social validation procedure was very simple and time efficient. Proctors spent approximately 90 minutes and instructors 45 minutes rating the 73 questions. Furthermore, neither group received training other than what was provided in the instructions for completing the questionnaire. Thus, with reasonable amounts of time and a simple procedure, useful and valuable task analyses were collected.

Conclusion

The present data suggest that conceptual levels in college course materials can be measured reliably. These data also support previous recommendations for the use of this method. It appears that educators and researchers can use this method due to its procedural reliability, ease of implementation, and small time requirements.

Reference Notes

1. Sanchez-Sosa, J.J., Semb, G., & Spencer, R. Using study guides to promote generalization performance in university instruction. Paper presented at the annual meeting of the American Psychological Association, Chicago, 1977.

References

Bloom, B.S. (ed.) Taxonomy of educational objectives: I - cognitive domain. New York: Longmans, Green, 1954.

Coldeway, D.O. Letter to the editor. Journal of Personalized Instruction, 1978, 3, 67-68.

Gagné, R.M. The conditions of learning (2nd ed.). New York: Holt, Rinehart, and Winston, 1970.

Gagné, R.M. Identifying objectives - task descriptions. In M.D. Merrill (ed.), Instructional design: Readings. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Kazdin, A.E. Assessing the clinical or applied importance of behavior change through social validation. Behavior Modification, 1977, 1, 427-452.

Keller, F.S. "Goodbye teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Semb, G., & Spencer, R. Beyond the level of recall: an analysis of complex educational tasks in college and university instruction. In L. Fraley and E. Vargas (Eds.), Behavior research and technology in college and university instruction. Gainsville, Florida: Department of Psychology, 1976.

Scheffé, H. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Winer, B.J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

Wolf, M.M. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis, 1978, 11, 203-214.

Table 1

The Mean Ratings of the SAEQ's by the Proctors and the Results of the Analysis of Variance

Dimensions

Difficulty	Importance	Thinking	Creativity
4.61±0.64	2.91±0.37	3.88±0.60	4.61±0.68

Summary Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Dimensions	141.74	3	47.25	202.69*
Questions	48.12	72	0.67	
Residual	50.35	216	0.23	
Total	240.17	291		

Note. mean ± S.D.

*p<.01.

Table 2

Pairwise Comparisons of Proctor Ratings of SAEQ's

Comparison	Degrees of Freedom	Scheffé F	Probability Value[a]
Difficulty vs. Importance	3, 216	451.11	<.01
Importance vs. Thinking	3, 216	148.54	<.01
Thinking vs. Creativity	3, 216	82.39	<.01
Difficulty vs. Creativity	3, 216	0.00	>.05
Importance vs. Creativity	3, 216	452.18	<.01
Difficulty vs. Thinking	3, 216	81.94	<.01

[a] $F'_{.01} = 11.64$

Table 3

The Mean Ratings of the SAEQ's by the Instructors and the Results of the Analysis of Variance

Dimensions

Difficulty	Importance	Thinking	Creativity
4.43±0.53	3.04±0.53	4.51±0.61	4.96±0.69

Summary Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Dimensions	151.05	3	50.35	282.48**
Questions	66.35	72	0.92	
Residual	38.50	216	0.18	
Total	255.90	291		

Note. mean ± S.D.

**p<.01.

Table 4

Pairwise Comparisons of Instructor Ratings of SAEQ's

Comparison	Degrees of Freedom	Scheffé F	Probability Value[a]
Difficulty vs. Importance	3, 216	405.52	<.01
Importance vs. Thinking	3, 216	441.31	<.01
Thinking vs. Creativity	3, 216	41.84	<.01
Difficulty vs. Creativity	3, 216	55.97	<.01
Importance vs. Creativity	3, 216	754.93	<.01
Difficulty vs. Thinking	3, 216	1.03	>.05

[a] $F'_{.01} = 11.64$

Table 5

Proctors' Versus Instructors' Ratings of SAEQ's

Dimension	Mean Rating by Instructors	Proctors	Degrees of Freedom	t separate	Probabilit Value
Difficulty	4.43±0.53	4.61±0.64	142.0	1.68	>.05
Importance	3.04±0.53	2.91±0.37	129.1	-1.74	>.05
Thinking	4.51±0.61	3.88±0.60	143.9	-6.23	<.01
Creativity	4.96±0.69	4.61±0.68	144.0	-3.11	<.01
Hotelling T^2 for differences among group means using all variables			4,141	211.22	<.01

Note. mean ± S.D.

Table 6

Pearson r Correlation Coefficients for the Four Dimensions As Rated By Proctors and Instructors

	Dimension			
	Difficulty	Importance	Thinking	Creativity
SAEQ	+.64**	+.22	+.59**	+.58**

*p<.05.

**p<.01.

Table 7

Intercorrelations of Objective Difficulty and Proctors' Ratings of SAEQ's

	Objective Difficulty	Difficulty	Importance	Thinking	Creativity
Objective Difficulty	+1.00				
Difficulty	+0.11	+1.00			
Importance	-0.03	0.00	+1.00		
Thinking	+0.07	+0.55**	+0.11	+1.00	
Creativity	+0.02	+0.35**	+0.05	+0.62**	+1.00

**p<.01.

Table 8

Intercorrelations of Objective Difficulty and Instructors' Ratings of SAEQ's

	Objective Difficulty	Difficulty	Importance	Thinking	Creativity
Objective Difficulty	+1.00				
Difficulty	+0.08	+1.00			
Importance	+0.05	+0.15	+1.00		
Thinking	+0.03	+0.86**	+0.12	+1.00	
Creativity	+0.03	+0.76**	+0.09	+0.86**	+1.00

**p<.01.

Table 9

Estimates of Reliability for Short Answer Essay Questions

	Dimensions			
	Difficulty	Importance	Thinking	Creativity
Instructors	+0.58**	+0.47**	+0.53**	+0.51**
Proctors	+0.89**	+0.68**	+0.86**	+0.87**

**p<.01.